

Assessing the Impact of Horizontal Error Correlations in Background Fields on Soil Moisture Estimation

ROLF H. REICHLE

*Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore, Baltimore, and Hydrological Sciences Branch,
NASA Goddard Space Flight Center, Greenbelt, Maryland*

RANDAL D. KOSTER

Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 13 March 2003, in final form 18 June 2003)

ABSTRACT

The importance of horizontal error correlations in background (i.e., model forecast) fields for large-scale soil moisture estimation is assessed by comparing the performance of one- and three-dimensional ensemble Kalman filters (EnKF) in a twin experiment. Over a domain centered on the U. S. Great Plains, gauge-based precipitation data is used to force the “true” model solution, and reanalysis data for the prior (or background) fields. The difference between the two precipitation datasets is thought to be representative of errors that might be encountered in a global land assimilation system. To ensure realistic conditions the synthetic observations of surface soil moisture match the spatiotemporal pattern and expected errors of retrievals from the Scanning Multichannel Microwave Radiometer (SMMR) on the *Nimbus-7* satellite. After filter calibration, average actual estimation errors in the (volumetric) root zone moisture content are $0.015 \text{ m}^3 \text{ m}^{-3}$ for the 3D-EnKF, $0.019 \text{ m}^3 \text{ m}^{-3}$ for the 1D-EnKF, and $0.036 \text{ m}^3 \text{ m}^{-3}$ without assimilation. Clearly, taking horizontal error correlations into account improves estimation accuracy. Soil moisture estimation errors in the 3D-EnKF are smallest for a correlation scale of 2° in model parameter and forcing errors, which coincides with the horizontal scale of difference fields between gauge-based and reanalysis precipitation. In this case the 3D-EnKF requires 1.6 times the computational effort of the 1D-EnKF, but this factor depends on the experiment setup.

1. Introduction

In a data assimilation system large-scale estimation of soil moisture relies on the merger of information from land surface models, meteorological forcing data, and satellite observations of land surface states. The key to optimal soil moisture estimation is the accurate representation of the error characteristics of each source of information. In large-scale assimilation systems computational considerations force us to assume that model, forcing, and observation errors are normally distributed and for the most part are characterized by their error covariances. Once these error covariances have been determined, data assimilation is a straightforward, albeit computationally demanding, application of standard optimization methods.

There are important differences between data assimilation in the atmosphere and at the land surface that are rooted in the nature of atmospheric and land surface

processes. Land surface models, as long as they do not include boundary layer processes, are basically just water- and energy-balance models, albeit highly complex and nonlinear. Perhaps the most important inputs to land surface models are meteorological forcing variables from screen-level observations or reanalysis data. Unlike atmospheric models, land surface models do not exhibit chaotic error growth. Consequently, errors in model parameters and time-dependent meteorological forcing inputs are much more important in land data assimilation than errors in initial conditions, which are typically of most interest in atmospheric data assimilation. Finally, land surface models typically describe the land surface as a collection of independent units or grid boxes. In the catchment land surface model of the NASA Seasonal-to-Interannual Prediction Project (NSIPP; Koster et al. 2000a), for example, the computational unit is the hydrological catchment, and all catchment-to-catchment fluxes such as groundwater or river flow are neglected.

It is tempting to construct the large-scale assimilation system as a collection of independent assimilation problems, one for each catchment or grid box. Such systems, sometimes known as one-dimensional (1D) assimilation

Corresponding author address: Rolf Reichle, NASA Goddard Space Flight Center, Code 900.3, Bldg 33, Rm A-110, Greenbelt Road, Greenbelt, MD 20771.
E-mail: reichle@janus.gsfc.nasa.gov

systems, are relatively easy to implement and computationally very affordable (Walker and Houser 2001; Reichle et al. 2002b; Crow and Wood 2003; Margulis et al. 2002). In these 1D assimilation systems, by construction only error correlations between state variables from the same computational unit are taken into account, and potential error correlations between states from different catchments or grid boxes are neglected.

The modeling of the land surface as a collection of independent catchments or grid boxes, however, does not imply that background (i.e., model forecast) errors of separate units are uncorrelated. Unmodeled fluxes across catchment boundaries such as river or groundwater flow are likely to generate error correlations in the model's predicted fields. Moreover, the approximation or neglect of true physical processes in the model can lead to horizontal correlations in model errors. For example, potential large-scale errors in model parameters such as vegetation inputs or soil hydraulic parameters may entail large-scale error features in model soil moisture. Finally and most importantly, errors in meteorological forcing data can be expected to exhibit large-scale correlations, leading to horizontal correlations in the errors of model states. If horizontal error correlations do in fact exist, taking them into account may be useful—they may allow the spreading of information from observed to unobserved catchments. The resulting impact on the accuracy of soil moisture estimation might be substantial.

Three-dimensional (3D) soil moisture assimilation methods (Reichle et al. 2001, 2002a) take into account error correlations of states that belong to different computational units. However, it has not yet been established whether 3D methods do in fact yield more accurate soil moisture estimates compared to 1D approaches. The answer to this question has wide implications for the design of emerging land data assimilation systems—the benefits of 3D methods must be traded off against their higher complexity and computational cost. In this paper, we assess the impact of horizontal error correlations in background fields on soil moisture estimation accuracy by comparing estimates from one- and three-dimensional versions of the ensemble Kalman filter.

The ensemble Kalman filter (EnKF) is a sequential data assimilation method that uses ensemble representations of the background (i.e., model forecast) error covariance (Evensen 1994). Many variants of the EnKF have been described in the literature. Among these are methods that use pairs of ensembles (Houtekamer and Mitchell 1998), hybrid approaches that combine ensembles with reduced-rank techniques (Heemink et al. 2001) or with variational methods (Hamill and Snyder 2000), particle filters (Pham 2001), and filters that avoid using random samples of observation error (Tippett et al. 2003). Because of its flexibility and robustness, the EnKF compares favorably to the extended Kalman filter and is particularly well suited to large-scale soil moisture estimation (Reichle et al. 2002b). In particular, the

EnKF requires no model adjoint or derivatives, can handle a wide range of model and forcing errors, and has been shown to yield accurate soil moisture estimates with few ensemble members. Moreover, a three-dimensional EnKF (3D-EnKF) for soil moisture estimation (Reichle et al. 2002a) was found to perform well against a variational assimilation method even though covariance localization techniques (see section 2b) were not used. Covariance localization methods have since been found to improve estimation accuracy, particularly for small ensemble sizes.

Here, to examine the importance of horizontal error correlations, we make use of the common “twin experiment” approach. That is, synthetic (model-generated) values of soil moisture are used to represent reality, which can then be considered perfectly known. “Satellite retrievals” of this reality are assimilated into a degraded version of the same modeling system. This degraded version of the system represents model errors by assigning different values to certain key model parameters and by using different precipitation forcing. Note that synthetic data are essential for this analysis. At this time we cannot perform the same study with in situ or satellite data, given that soil moisture observations of sufficient density and quality are simply not available across the spatial scales we consider—we would have no way of determining what “reality” is, so we would have no way of quantifying any improvement in performance associated with the consideration of horizontal error correlations.

The twin experiment presented here has important implications for the design of emerging land assimilation systems. It serves to isolate and quantify, for the first time ever, the information added to a land data assimilation system by the consideration of horizontal error correlations. Presumably a similar improvement would be felt when satellite soil moisture data are assimilated. The existence of surface soil moisture data retrieved from the C-band Scanning Multifrequency Microwave Radiometer (SMMR) instrument for the period 1979–87 (Owe et al. 2001), the imminent availability of similar soil moisture data from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) instrument, and the eventual availability of passive L-band sensor data make the consideration and evaluation of such distributed data assimilation approaches a timely problem.

2. The ensemble Kalman filter

The standard Kalman filter is the optimal sequential data assimilation method for linear dynamics and measurement processes with Gaussian error statistics. The ensemble Kalman filter is a Monte Carlo approach to the nonlinear filtering problem (Evensen 1994) and is based on the approximation of the conditional probability densities of interest by a finite number of randomly generated model trajectories. In this section we provide

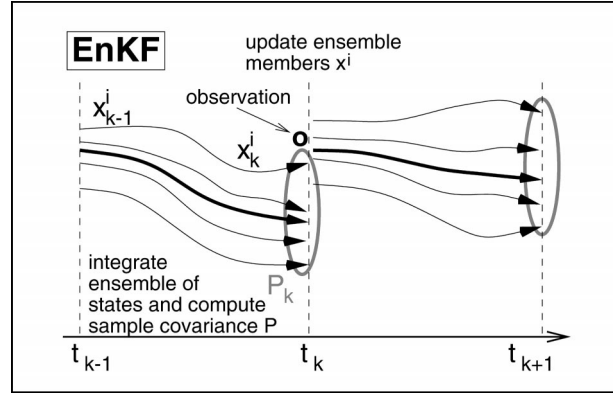


FIG. 1. Schematic of the ensemble Kalman filter (EnKF).

a brief summary of the EnKF. A comprehensive discussion of the EnKF equations can be found in (Reichle et al. 2002b).

We start by collecting the relevant model prognostic variables from all catchments into the state vector \mathbf{x} . Figure 1 depicts how the EnKF works sequentially from measurement time $k - 1$ to the next measurement time k , applying in turn a forecast step and an update step. During the forecast step the EnKF propagates an ensemble of state vectors \mathbf{x}^i using a corresponding ensemble of N random realizations of model error fields \mathbf{w}^i .

$$\mathbf{x}_k^{i-} = \mathbf{f}_k(\mathbf{x}_{k-1}^{i+}) + \mathbf{w}_k^i \quad i = 1, 2, \dots, N. \quad (1)$$

Here, the nonlinear operator $\mathbf{f}_k(\cdot)$ includes all deterministic forcing data (e.g., observed rainfall). The superscripts $-$ and $+$ refer to states and covariances before and after updates, respectively. Each state vector represents a particular realization of the possible model trajectories with certain random errors in model parameters and/or a particular set of errors in forcing. Such uncertainties are summarized in the model error term \mathbf{w}_k , which is assumed to be a zero mean random variable with covariance \mathbf{Q}_k . Since deterministic forcing data are an integral part of the model, errors in the forcing are sometimes also referred to as model errors.

With the observations \mathbf{y}_k available at time k , the state \mathbf{x}_k^{i-} of each ensemble member is updated to a new value \mathbf{x}_k^{i+} :

$$\mathbf{x}_k^{i+} = \mathbf{x}_k^{i-} + \mathbf{K}_k(\mathbf{y}_k^i - \mathbf{H}_k \mathbf{x}_k^{i-}) \quad i = 1, 2, \dots, N \quad (2)$$

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1}. \quad (3)$$

Simply put, the Kalman gain matrix \mathbf{K}_k represents the relative weights given to the model forecast \mathbf{x}_k^{i-} and the observations \mathbf{y}_k . The state (or background) error covariance \mathbf{P}_k is computed as the sample covariance of the ensemble (\mathbf{x}_k^{i-} , $i = 1, 2, \dots, N$) prior to the update. The operator \mathbf{H}_k relates the states to the measured variables in the measurement equation $\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$, where \mathbf{v}_k reflects measurement instrument errors and errors of representativeness (assumed zero mean with covariance \mathbf{R}_k). Note that in the traditional EnKF update

the observations are perturbed by adding a random realization \mathbf{v}_k^i of the measurement error such that $\mathbf{y}_k^i = \mathbf{y}_k + \mathbf{v}_k^i$ (Burgers et al. 1998). The reduction of the uncertainty resulting from the update is reflected in the reduction of the ensemble spread (Fig. 1). The EnKF state estimate at all times is simply the mean of the ensemble members.

a. One- and three-dimensional EnKF

To illustrate the concept of horizontal error propagation, let us assume that the domain contains only two catchments, a and b , with one scalar state each, and that ρ denotes the covariance of the state forecast errors in the two catchments:

$$\mathbf{x}_k = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \quad \mathbf{P}_k = \begin{pmatrix} \sigma_a^2 & \rho \\ \rho & \sigma_b^2 \end{pmatrix}. \quad (4)$$

Let us further assume that the state of catchment a is observed directly, and that the state of catchment b is not observed at all; that is, $\mathbf{y}_k = y_a$ and $\mathbf{H}_k = [1 \ 0]$ with $\mathbf{R}_k = \sigma_v^2$. Dropping the ensemble index i and the time index k for clarity, the resulting update of the model forecast by the observation y_a is readily obtained from (2) and (3):

$$\begin{pmatrix} x_a^+ - x_a^- \\ x_b^+ - x_b^- \end{pmatrix} = \frac{(y_a - x_a^-) \begin{pmatrix} \sigma_a^2 \\ \rho \end{pmatrix}}{(\sigma_a^2 + \sigma_v^2)} \quad (5)$$

In this simple example, the state of the unobserved catchment b is updated from an observation y_a in catchment a based on the error covariance ρ between the two catchments. In other words, some of the information in the observed catchment a is transferred into the unobserved catchment b , provided the error correlation between the states in the two catchments is nonzero. The importance of this transfer of information is the focus of this paper.

Since catchments are modeled independently, the forecast equation (1) implies that error correlations between states from different catchments (as reflected in \mathbf{P}_k) exist only if generated by appropriate off-diagonal elements in the model error covariance \mathbf{Q}_k or because such correlations were already present in the updated ensemble at the previous time $k - 1$ (as reflected in \mathbf{P}_{k-1}^+). In the one-dimensional EnKF (or 1D-EnKF) all correlations between different catchments are neglected in the model error covariance \mathbf{Q}_k , the measurement error covariance \mathbf{R}_k , and in the initial state error covariance \mathbf{P}_0 . Under these assumptions the state (or background) error covariance matrix \mathbf{P}_k always maintains a block-diagonal structure, with each block corresponding to a different catchment, and the update is computed independently for each catchment. In other words, the 1D-EnKF is applied to each catchment independently, which offers great computational savings but neglects potential horizontal propagation of error information.

If, however, horizontal correlations exist between

model, forcing, or observation errors of different catchments, these inevitably translate into nonzero off-diagonal elements of the state error covariance matrix \mathbf{P}_k^- , and the update ought to be computed for all catchments simultaneously. For large domains such a global update is not computationally feasible because of the large size of the matrices involved. Fortunately, a global update is also not necessary. A reasonable scale for horizontal error correlations in continental or global soil moisture fields is much smaller than the domain size. In this case, covariance localization (or compact support) techniques in combination with a parallel implementation make it possible to take medium-range horizontal error correlations into account within the practical constraints of limited computational resources.

b. Covariance localization

The modest ensemble size that can be afforded with large-scale EnKF systems introduces spurious long-range correlations in the forecast ensemble. This statistical noise drowns out any actual weak correlations that may arise at larger separation distances and makes their accurate estimation impossible (Houtekamer and Mitchell 2001; Hamill et al. 2001). It is therefore better to suppress correlations beyond a certain separation distance by using a Hadamard (or Schur) product with a local compactly supported correlation function. This product is applied to the sample covariance terms $\mathbf{P}_k^- \mathbf{H}_k^T$ and $\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T$ in (3). The compactly supported correlation function is taken from [Gaspari and Cohn 1999, their Eq. (4.10)]. A formal discussion of the modified EnKF equations can be found in (Keppenne and Rienecker 2002). The covariance localization greatly improves the conditioning of the key matrices in the update equation (2).

c. Parallelization

Although the covariance localization imposes a banded structure on the forecast state error covariance matrix, the dimension of the global matrix that must be inverted in the update equation (3) is unchanged and too large to be affordable. An efficient implementation of the 3D-EnKF must rely on parallelization. Since the land surface model is inherently parallel, the EnKF forecast step is trivially parallelized, and only the update step deserves a brief discussion. The basic idea is to decompose the computational domain into smaller subdomains, each of which is updated independently. Since the sample covariances are localized by applying the Hadamard product (section 2b), only observations within compact support distance from each subdomain can influence the update of states in the subdomain. The global update is thus replaced with a simultaneous update of the smaller subregions that can be carried out in parallel. A formal discussion of the parallelization can be found in (Hou-

tekamer and Mitchell 2001; Keppenne and Rienecker 2002).

3. Land model and experiment setup

a. Land surface model

For our twin experiment we use the NSIPP catchment land surface model (Koster et al. 2000a; Ducharme et al. 2000), which uses the hydrological catchment rather than a rectangular grid cell as the computational unit. Its viability has been demonstrated in a series of land model intercomparison experiments (Bowling et al. 2003; Boone et al. 2004). Recall that catchments are independent in the model (but coupled in the 3D-EnKF, section 2a). In each catchment, three prognostic variables related to soil water allow the diagnostic calculation of horizontal soil moisture variability, which in turn allows for conceptually improved treatments of evaporation and runoff. The first prognostic variable, the catchment deficit, determines the equilibrium soil moisture profile and is defined as the amount of water needed to bring the entire catchment to saturation. To allow for nonequilibrium vertical distribution of water, two additional variables (surface and root zone excess) describe deviations from the equilibrium profile in the surface and root zone layers.

From the prognostic catchment deficit, root zone excess, and surface excess, we can diagnose average soil moisture content in the 2-cm surface layer, the 1-m root zone layer, and the total soil profile (Walker and Houser 2001). We refer to these diagnostic variables as surface, root zone, and profile soil moisture, respectively. In total there are 22 prognostic variables per catchment (3 for soil moisture, 1 for canopy interception, 9 for soil and canopy temperatures, and 9 for snow). In our Kalman filtering application, however, we use only the three model prognostic variables that are directly related to soil moisture as state variables for the assimilation (catchment deficit, root zone excess, and surface excess). We assimilate synthetic observations of the (diagnostic) surface soil moisture.

b. Twin experiment

Our twin experiment is conducted over the Great Plains region of the United States. The computational domain is composed of all catchments that are fully contained within 31°–48°N and 110°–90°W. The 731 catchments in the domain have an average area of about 4000 km², and thus the spatial resolution is approximately 0.5°. The domain was chosen partly because high-quality precipitation data are available there and partly because soil moisture memory in the region may be a source of seasonal predictability of summer rainfall (Koster et al. 2000b; Koster and Suarez 2003).

The twin experiment starts with a model integration that serves as the “true” solution and is meant to rep-

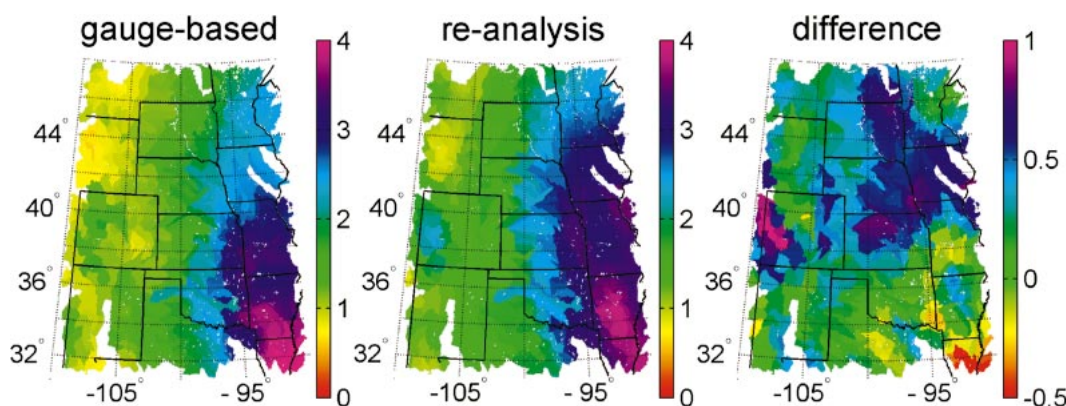


FIG. 2. Mean total precipitation 1983–86 (mm day^{-1}): (left) gauge-based, (middle) reanalysis, and (right) difference (reanalysis minus gauge-based).

resent nature. We start from a spinup initial condition on 1 January 1983 and integrate the model until 31 December 1986 using standard model parameters and high-quality precipitation data from Higgins et al. (2000). These precipitation data are available daily from 1948 to 1998 on a 0.25° grid covering the continental United States and are based on about 15 000 gauges on a typical day. All other meteorological forcing data (including, e.g., air temperature and humidity) are taken from the 15-yr European Centre for Medium-Range Weather Forecasting Reanalysis (ERA-15; Gibson et al. 1997) which is available from 1979 to 1993 with an effective horizontal resolution of about 1.6° . The spinup initial condition was derived by repeatedly integrating the model for 10 yr with 1979 forcing and subsequent integration from 1 January 1980 to 1 January 1983. Monthly leaf area index and greenness fraction vary interannually and were derived from data collected by the Advanced Very High Resolution Radiometers (AVHRRs; Guillevic et al. 2002).

Additional integrations for the same period with errors introduced in the model forcing and parameters are then compared to the true integration. In one of these integrations, the “prior” integration, surface soil moisture is not assimilated. In further (“assimilation”) integrations, surface soil moisture is assimilated with either the 1D- or 3D-EnKF. Forcing errors are imposed by replacing the gauge-based precipitation with reanalysis precipitation from ERA-15. Figure 2 illustrates the two precipitation datasets and their difference. There is a strong precipitation gradient from west to east across the domain, with the highest precipitation recorded in the southeast in both datasets. The mean reanalysis precipitation is higher than the gauge-based in the northern half of the domain and less in the far southeastern corner. Imposed parameter errors include the replacement of the timescale parameters for moisture flow between the surface excess, root zone excess, and catchment deficit. Specifically, we use timescale parameters that were derived for a 5-cm surface layer and a vertical decay factor $\gamma = 3.26$ for the saturated hydraulic conductivity with

depth (rather than for the 2-cm layer and $\gamma = 2.17$ that we use in the true integration). Finally, the monthly climatology from 1982 to 1990 is used for the leaf area index and greenness fraction instead of interannually varying parameters. The initial condition for 1 January 1983 is generated by repeating the spinup procedure described earlier with the changed forcing data and model parameters. Collectively, these “wrong” forcing and parameter inputs represent our imperfect knowledge of the true land processes.

Synthetic observations of soil moisture content in the 2-cm surface layer (“surface soil moisture”) for the 1D and 3D filters are derived from the true fields by adding random measurement noise. The spatial and temporal distribution of the synthetic observations must be chosen with care, because the importance of horizontal error correlations obviously depends on the spatial and temporal distribution of the assimilated observations. The biggest advantage of taking horizontal error correlations into account is that information from observed catchments may be spread to unobserved locations. If all catchments of the domain were observed simultaneously, it is unlikely that horizontal error correlations would have a strong impact. In practice, all catchments are not observed simultaneously because a satellite scans the globe in swaths and because the properties of some catchments, notably dense vegetation, defy the retrieval of soil moisture from satellite brightness temperatures. To ensure the most realistic experiment conditions, our synthetic observations follow the spatial and temporal observation pattern of the Scanning Multichannel Microwave Radiometer on the *Nimbus-7* satellite.

SMMR is a passive microwave radiometer with five channels ranging from 6.6 GHz (C-band) to 37 GHz. Overpasses of a given location are about once every 3 days on average. Soil moisture retrieval errors primarily depend on the vegetation density at the time of the measurement. In catchments with dense vegetation (vegetation optical depth greater than 0.65) soil moisture cannot be accurately retrieved from C-band observations. For lightly vegetated catchments, observation error stan-

standard deviations range from 0.046 to 0.095 $\text{m}^3 \text{m}^{-3}$ (De Jeu 2003). We use these observation error standard deviations when adding synthetic noise to the synthetic soil moisture observations and also in the EnKF algorithm. Note that observation errors for C-band soil moisture retrievals are very large relative to the dynamic range of soil moisture, which is typically less than 0.35 $\text{m}^3 \text{m}^{-3}$. For simplicity, we neglect potential horizontal error correlations in the observations even though they are probably present in satellite data. Our analysis focuses on horizontal correlations in model forcing and parameter errors. Taking horizontal error correlations of the satellite data into account would presumably further improve the relative performance of the 3D-EnKF.

c. Filter calibration

The setup of the twin experiment implies that we do not know the exact statistics of the model and forcing errors. But filter performance depends strongly on our choice of model error parameters, so we must choose them very carefully. It is quite possible, for example, that the 1D-EnKF can compensate for neglecting horizontal error correlations through increased model error standard deviations. To ensure a fair comparison of the 1D- and the 3D-EnKF, we find the parameters that allow each filter to perform the best it can.

In data assimilation integrations, the ensemble spread is generated and maintained by adding perturbations to model fields and forcing inputs. Besides perturbing the precipitation forcing of each ensemble member with a lognormal distributed multiplicative error, we add synthetic model error fields to the three state variables related to soil moisture (catchment deficit, root zone, and surface excess). For the 1D-EnKF we generate spatially uncorrelated errors directly in catchment space. The correlated error fields required for the 3D-EnKF are generated on a 0.5° grid with a fast Fourier transform algorithm and subsequently interpolated to catchment space. In all cases we assume that the standard deviation of each type of model error is identical for all catchments and use a normal probability distribution. Errors for different variables are assumed uncorrelated. Finally, the time series of error fields is modeled as an autoregressive process of order one with a correlation time of 1 day.

During the calibration of the model error variances we set the e -folding scale of all horizontal error correlations (including forcing perturbations) in the 3D-EnKF to 2° in latitude/longitude coordinates. We also impose a covariance localization scale of 5° , that is all error correlations beyond 5° separation distance are neglected (section 2b). The 2° error correlation scale matches the scale of the difference fields between gauge-based and reanalysis precipitation (determined independently). It is possible that simultaneous calibration of error variances and horizontal scale might further improve performance of the 3D-EnKF, but it is too com-

putationally expensive and therefore avoided. With all inputs fixed except the magnitude of the model and forcing error variances (for surface excess, root zone excess, catchment deficit, and precipitation), we calibrate these remaining parameters to achieve the best possible filter performance. Since the twin experiment is designed such that the true solution is known, a convenient measure of estimation performance is the actual error, which is the difference between the true soil moisture and its EnKF estimate. As an aggregate measure of filter performance we choose the average actual errors in the root zone soil moisture, where the average is taken in the root-mean-square sense over all catchments from 1 January 1983 to 31 December 1986. We choose errors in root zone soil moisture because its memory may be important for seasonal prediction.

For each filter, we computed the aggregate estimation errors of about 100 integrations with different model error variances. The results of a previous calibration experiment of the 1D-EnKF (Reichle et al. 2002b) provided helpful guidance in the selection of trial model error variances. For instance, the model error variance in surface excess matters little because we also perturb the precipitation inputs. The set of model and forcing error variances that yields minimum root zone soil moisture errors turns out to be the same for the 1D- and the 3D-EnKF. This implies that at least within our coarse calibration the 1D-EnKF does not compensate for its lack of horizontal error correlations by requiring different input error standard deviations than the 3D-EnKF. The calibrated error standard deviations for both filters are 0.72 mm day^{-1} for the surface excess, 0.072 mm day^{-1} for the root zone excess, 0.72 mm day^{-1} for the catchment deficit, and 50% of magnitude for precipitation errors. The resulting parameters are almost identical when we calibrate against the average of errors in surface or profile moisture contents. Our calibration of model error parameters serves mainly to make a fair comparison of the 1D- and 3D-EnKF possible. There are many adaptive methods to determine error statistics during filter operation (Dee 1995).

4. Results and discussion

a. Soil moisture estimates

Figure 3 shows time-average estimation errors for surface, root zone, and profile soil moisture without assimilation (prior) and for the calibrated filters. Without assimilation the largest errors are concentrated in the eastern and northern part of the domain. This is consistent with the geographical distribution of differences between the gauge-based and reanalysis precipitation (Fig. 2). In the southwestern part of the domain, differences between gauge-based and reanalysis precipitation are small and prior soil moisture already agrees well with the truth even without assimilation. Where prior estimation errors are high, both filters achieve a

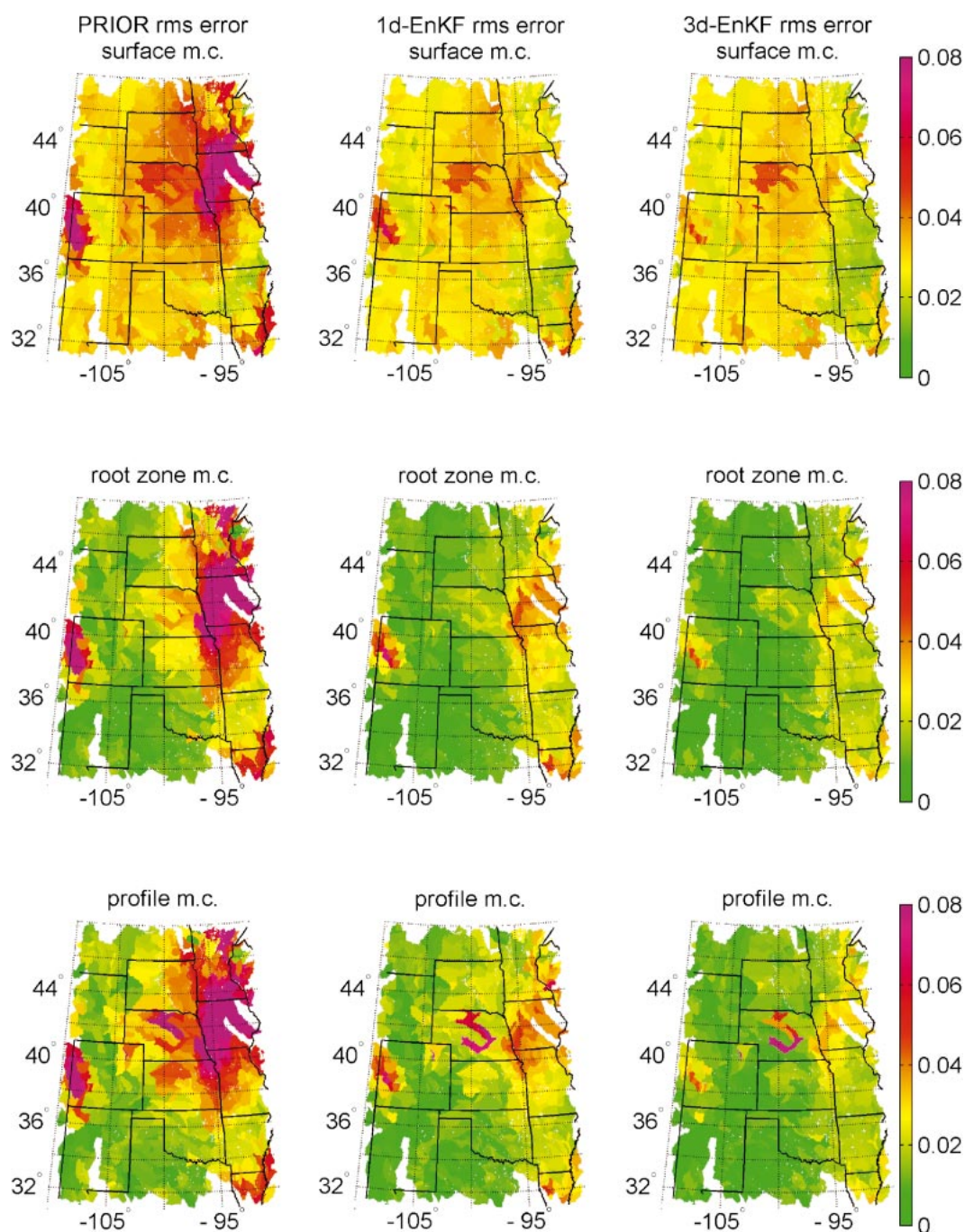


FIG. 3. Time-average error of the moisture content (m.c.) (left) prior to the assimilation, (middle) for the 1D-EnKF, and (right) for the 3D-EnKF. Shown are errors for (top) surface, (middle) root zone, and (bottom) profile soil moisture content. The average is from 1 Jan 1983 to 31 Dec 1986 in the root-mean-square (rms) sense. Units are volumetric soil moisture ($\text{m}^3 \text{m}^{-3}$).

significant reduction of errors. Overall the errors of the 3D-EnKF estimates show the smoothest and most homogeneous structure of all estimates. The most interesting feature in Fig. 3 is the reduction of the large prior errors in the eastern half of the domain and on the western boundary by the two filters. In these subregions the

1D-EnKF improves significantly on the prior estimates, and the 3D-EnKF yields a further reduction of the errors.

The time- and domain-average estimation errors are listed in Table 1. Average root zone soil moisture errors in the 3D-EnKF decrease by 58% from their prior values compared to a decrease of just 47% in the 1D-EnKF

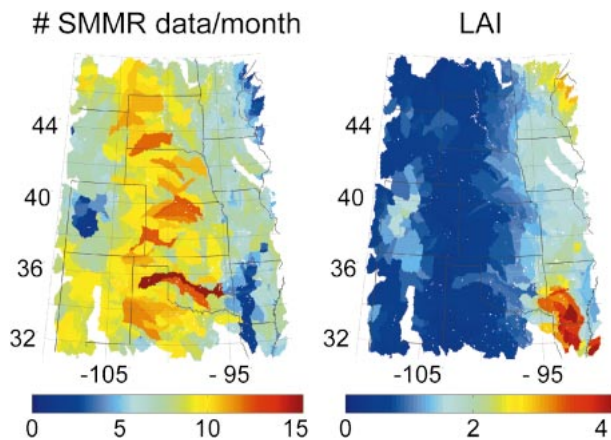


FIG. 4. (left) Average number of synthetic SMMR data of surface soil moisture that were assimilated per month (1983–86). (right) Average leaf area index (LAI), 1983–86.

(from 0.036 to 0.015 $\text{m}^3 \text{m}^{-3}$ in the 3D-EnKF and to 0.019 $\text{m}^3 \text{m}^{-3}$ in the 1D-EnKF). Almost identical relative decreases can be seen for errors in the profile moisture content. Similarly, domain-average evaporation errors in the 3D-EnKF (1D-EnKF) decrease by 17% (8%) from their prior values. In terms of the surface soil moisture the 1D- and 3D-EnKF exhibit almost identical performance and show a clear improvement over prior errors (from 0.040 to 0.029 $\text{m}^3 \text{m}^{-3}$ in the 3D-EnKF and to 0.030 $\text{m}^3 \text{m}^{-3}$ in the 1D-EnKF). For runoff, there is only a slight improvement in both filters, and for most of the calibration experiments runoff estimates from the EnKF are in fact worse than prior estimates. This is to be expected because runoff is not a state variable and is not updated in the filter. While the soil moisture update obviously improves soil moisture, runoff is a nonlinear function of precipitation, which is not updated. In fact, the very nature of precipitation perturbations implies that more extreme rain events are present in a few ensemble members, in particular for strong (high variance) precipitation perturbations. This typically leads to an overestimation of runoff in the filter

TABLE 1. Rmse of moisture content (m.c.), total evaporation, and runoff for the prior (no assimilation) and the calibrated filters.

	Prior	1D-EnKF	3D-EnKF
Surface m.c. ($100 \text{ m}^3 \text{m}^{-3}$)	4.0	3.0	2.9
Root zone m.c. ($100 \text{ m}^3 \text{m}^{-3}$)	3.6	1.9	1.5
Profile m.c. ($100 \text{ m}^3 \text{m}^{-3}$)	4.0	2.2	1.7
Evaporation (mm day^{-1})	1.2	1.1	1.0
Runoff (mm day^{-1})	3.0	2.8	2.8

relative to the prior. In summary, runoff estimates from the filter should be used with caution.

b. Data volume and filter performance

Data volume is critical to the performance of the filter, and it is also critical to the difference in the performances of the 1D- and the 3D-EnKF. The superior performance of the 3D-EnKF can be understood by analyzing the average number of surface soil moisture observations that were assimilated per month. Figure 4 shows that the volume of SMMR soil moisture retrievals is typically small where the leaf area index (LAI) is large. The areas where the 3D-EnKF shows superior performance are also areas of relatively low SMMR data volume. In this context the superior performance of the 3D-EnKF can be explained by its ability to spread information from observed to unobserved catchments. Note that because of details in the grid-to-catchment interpolation, the data volume also depends weakly on the size of the catchment. Fortunately, due to the synthetic nature of the experiment, this does not affect our results.

Figure 5 breaks down the error information by catchments and bins according to SMMR data volume. The scatterplots show time-average rmse of root zone soil moisture for the 1D-EnKF versus the 3D-EnKF. When there are fewer than five SMMR observations per month available, errors are generally large and the 3D-EnKF does significantly better than the 1D-EnKF. For catchments with higher observation volume, errors become smaller, and the difference in performance between the

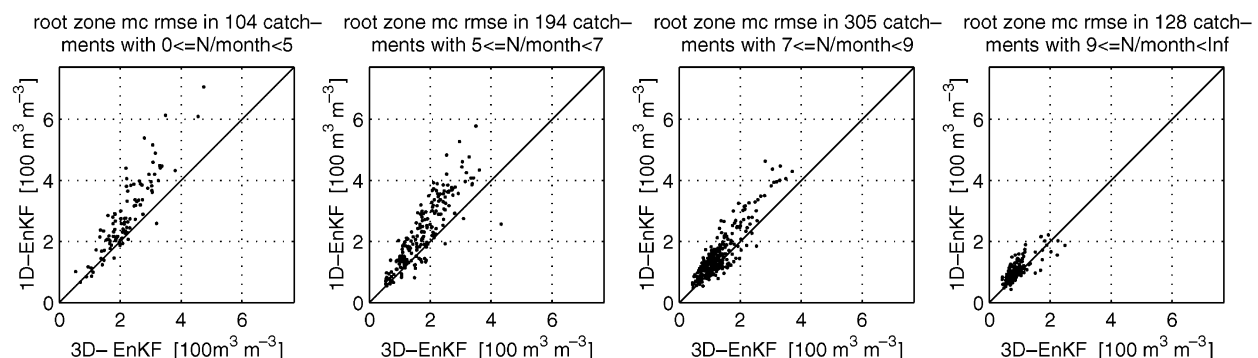


FIG. 5. Time-average rmse of root zone soil moisture of each catchment for the 1D-EnKF vs same for the 3D-EnKF, broken down by avg number of SMMR data available per month.

1D- and the 3D-EnKF gradually decreases. For catchments with more than nine observations per month there is no difference in performance between the 1D- and 3D-EnKF. These findings are consistent with results by Hamill and Snyder (2000) showing that a hybrid 3D-EnKF-variational atmospheric assimilation scheme is superior to a simple variational method [three-dimensional variational data assimilation (3DVAR)] primarily in data-sparse regions.

The filter calibration (section 3c) was carried out with a SMMR observation pattern that was subject to only marginal quality control. When actual SMMR retrievals are assimilated, additional quality control will be required, and the data volume will decrease relative to our baseline experiments. When the volume of SMMR data decreases, performance differences between the 1D- and 3D-EnKF will likely increase. To explore further the impact of data volume on the relative performance of the filters, we conducted a suite of supplemental experiments in which the number of assimilated data were gradually reduced in two ways. In the first strategy, (synthetic) data were withheld from the assimilation in an entirely random pattern. In the second, data were withheld mimicking a “quality control” strategy.

In the quality control strategy, data were withheld according to their observation time and quality. Observation times for SMMR are local noon and midnight. In one set of experiments, only daytime data were assimilated, in another only nighttime data were used, and in a third both daytime and nighttime data were assimilated. Moreover, data were withheld according to the retrievals of vegetation optical depth from the actual SMMR observations (De Jeu 2003). We assimilated data only if the retrieved optical depth was less than a specified value (namely, ∞ , 0.55, 0.5, 0.45, or 0.4). This strategy yielded a total of 15 experiments, covering all combinations of the three time-of-day choices and the five specified optical depth thresholds. Withholding data in this way mimics the potential effects of reasonable quality control measures. Nighttime soil moisture retrievals are typically more accurate than daytime retrievals because the temperature of the soil and canopy system is more homogeneous at night and thereby more easily retrieved from the 37-GHz channel. Similarly, soil moisture retrievals in less-vegetated catchments are more trustworthy. All experiments cover the entire domain from January 1983 to December 1986.

Figure 6 shows the time- and domain-average rmse of root zone soil moisture plotted against relative data volume. Relative data volume equal to zero corresponds to the prior (no assimilation) integration, and relative data volume equal to one corresponds to assimilating all synthetic data that were used in the filter calibration. As the volume of assimilated data increases, errors decrease. The decrease is steep for small data volumes and flattens out as the data volume increases. For all possible withholding strategies, the 3D-EnKF performs better than the 1D-EnKF. The performance of the two filters

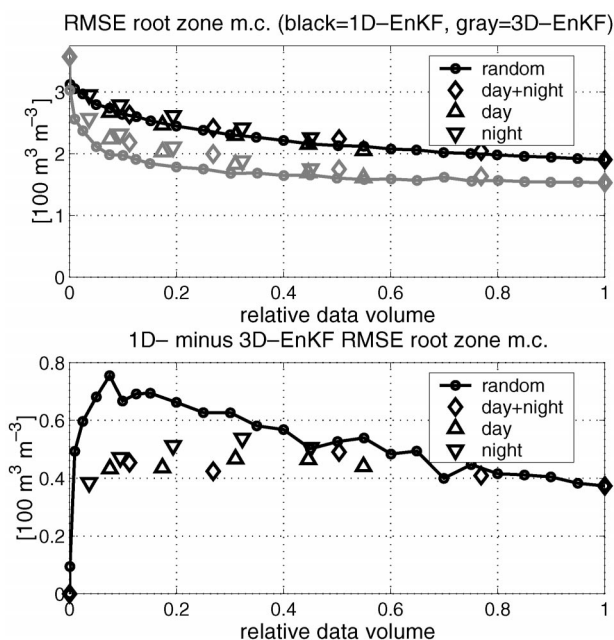


FIG. 6. Time- and domain-average rmse of root zone soil moisture vs relative data volume. Data volume is varied by withholding data randomly (circles/solid line) or by withholding data according to their SMMR-retrieved vegetation optical depth. In the latter case, we use either daytime data only (triangles pointing up), nighttime data only (triangles pointing down), or both daytime and nighttime data (diamonds).

is by definition the same as the data volume approaches zero (no assimilation). But even for very small but non-zero data volumes the 3D-EnKF does significantly better than the 1D-EnKF. There appears to be a maximum performance advantage of the 3D-EnKF somewhere between no data and full SMMR data coverage. Its exact location depends on the withholding strategy and is difficult to pin down because of statistical noise. Note that the random withholding strategy places the maximum very close to zero, while the quality control strategy places it closer to 0.5. This is consistent with expectations, since the random withholding strategy plays better into the hands of the 3D-EnKF's ability to spread information from observed to unobserved locations.

c. Annual cycle and horizontal scales

Soil moisture estimation errors are not constant throughout the year. To determine a crude climatology, we repeated the twin experiment for the extended period from 1 January 1979 to 31 August 1987 using the calibrated model and forcing error parameters. Figure 7 shows estimation errors in root zone soil moisture for each month averaged over the extended experiment period. The seasonality of the prior soil moisture errors reflects the seasonality of precipitation and its errors. Prior soil moisture errors are largest in spring and early summer, precisely when a 3-month seasonal forecast of

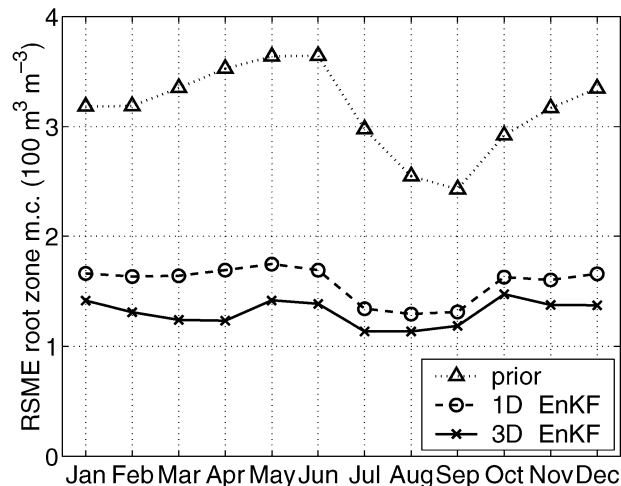


FIG. 7. Annual cycle of domain-average rmse of root zone soil moisture.

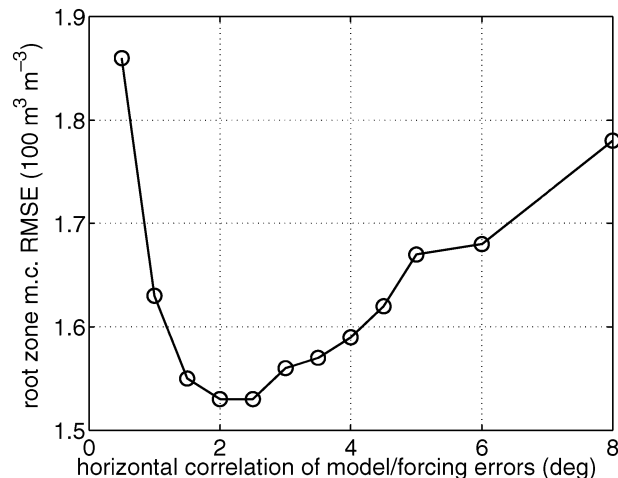


FIG. 8. Time- and domain-average rmse of root zone soil moisture for varying inputs of horizontal error correlation scale to the 3D-EnKF.

summer precipitation would be initialized. The seasonality of 1D- and 3D-EnKF soil moisture errors reflects again the seasonality of precipitation and its errors, but is also affected by the seasonality of SMMR data availability. Filter estimation errors do not exhibit the seasonal peak of the prior errors in late spring and early summer. This is due to the increased availability of soil moisture retrievals in spring, when soils are no longer frozen and vegetation is not yet too dense. The filters are thus able to maintain roughly constant performance when the prior (no assimilation) errors increase significantly. Interestingly, the performance advantage for the 3D-EnKF is also largest in spring and early summer.

The filter calibration discussed in section 3c involved only the calibration of model and forcing error standard deviations while the horizontal error correlation scale was fixed at 2° . We carried out another series of 3D-EnKF experiments imposing the calibrated error standard deviations but using a range of different e -folding scales for the horizontal error correlations. In these experiments the covariance localization scale was set to 2.5 times the error correlation (e -folding) scale. Each experiment covered the entire domain from January 1983 to December 1986. Figure 8 shows 3D-EnKF estimation errors in root zone soil moisture as a function of the horizontal error correlation scale. As the horizontal scale increases from zero, soil moisture estimation errors decrease and reach a minimum at about 2° . This optimal horizontal scale is consistent with the horizontal scale of the difference fields between gauge-based (true) and reanalysis (prior) precipitation data, as determined independently. The fact that these scales match corroborates the proper operation of the 3D-EnKF.

The workings of the 3D-EnKF can be further illustrated by examining the marginal gain for a unit innovation of a single observation. We define a unit innovation as a difference of $0.1 \text{ m}^3 \text{ m}^{-3}$ between a hypo-

thetical surface soil moisture observation and the corresponding model forecast. When such a hypothetical unit innovation is assimilated by itself, the resulting soil moisture increments are called the marginal gain. If many observations are assimilated simultaneously, the marginal gain of one of the observations is precisely the column of the Kalman gain matrix that corresponds to this observation (up to a factor). The total increment results as the sum of the columns of the Kalman gain matrix multiplied by the individual innovations [Eq. (2)].

Figure 9 shows the marginal gain for a unit innovation in a single catchment in the northeastern corner of the domain at 1800 UTC 20 June 1985 for the 3D-EnKF. The marginal gain is strongest in and very near the catchment where the observation is taken. Catchments within a radius of a few degrees from the observed catchment are also updated. Even for the observed catchment, the size of the increment is small. For a unit innovation of $0.1 \text{ m}^3 \text{ m}^{-3}$, the increment of the observed catchment is only on the order of $0.01 \text{ m}^3 \text{ m}^{-3}$ for the surface moisture content, and $0.001 \text{ m}^3 \text{ m}^{-3}$ for the root zone moisture and profile moisture contents. The increments are small because the observation error of soil moisture retrievals from C-band brightness data is large relative to the dynamic range of soil moisture. To compensate for such large observation errors the filter is tuned (or calibrated) such that it produces only small increments. The cumulative effect of many small increments is what eventually gets the filter estimate close to the truth.

The structure of the marginal gain in Fig. 9 appears complex mainly for three reasons. First, soil moisture error standard deviations vary strongly with soil moisture content, and soil moisture fields are not typically smooth. Second, the resolution of the irregular catchments is coarse compared to the correlation scales. Third, there is residual noise due to the small ensemble

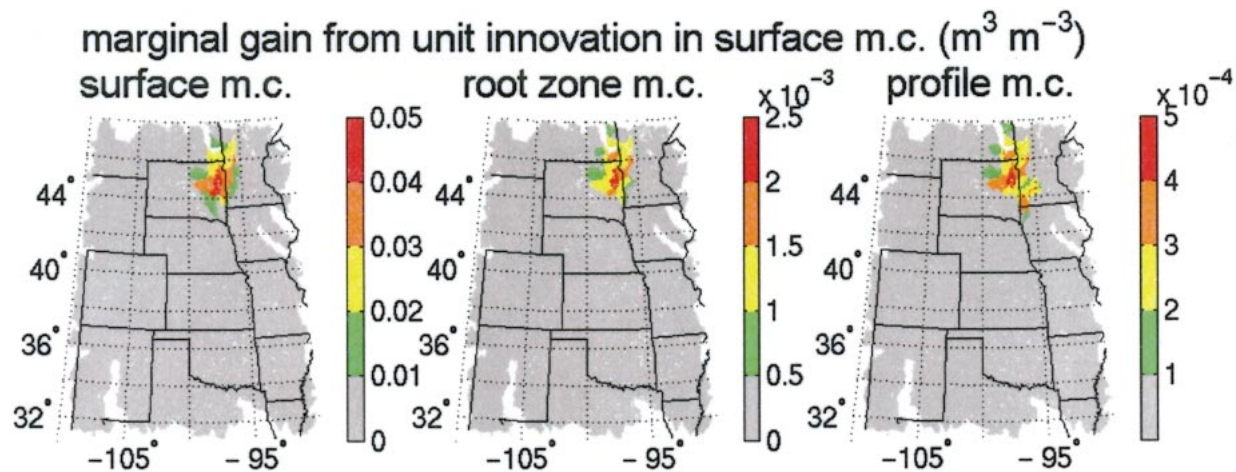


FIG. 9. Incremental update (or marginal gain) of the 3D-EnKF resulting from a $0.1 \text{ m}^3 \text{ m}^{-3}$ innovation 1800 UTC 20 Jun 1985 in a single catchment.

size. In comparison, the structure of the 1D-EnKF marginal increments (not shown) is trivial because only the observed catchment is updated and has a nonzero marginal gain. Due to residual noise, the distributed marginal gain of the 3D-EnKF may not be optimal and other approaches, in particular a hybrid 3D-EnKF-variational scheme (Hamill and Snyder 2000), may have added benefits. It is nevertheless obvious from the results discussed earlier that the distributed update of the 3D-EnKF is superior to the local nature of the 1D-EnKF in our experiment.

d. Ensemble size and computational demand

A great deal of optimism is perhaps necessary when the EnKF is used with just a few ensemble members in problems having large state vector dimensions (order of 10^3 in our case, easily exceeding 10^5 in global applications). With the small ensemble size, only a very small subspace of the error space is sampled, and any updates are necessarily restricted to this subspace. Moreover, the finite number of ensemble members in the EnKF results in undesirable statistical (or sampling) noise in addition to residual errors that are not due to the finite ensemble size. Finally, ensemble size is intimately related to computational demand. In this section we investigate the impact of ensemble size on estimation errors and take a closer look at the computational demands of both filters.

Table 2 shows the convergence of estimation errors in root zone soil moisture when the ensemble size is

increased from 4 to 128 members. When compared to the differences between the 1D- and 3D-EnKF, the errors across the range of ensemble sizes for each filter are fairly close. Even the 3D-EnKF with just 4 ensemble members is still more accurate than the 1D-EnKF with 128 ensemble members. In this sense our findings about the relative performance of the 1D- and 3D-EnKF are robust, and our results do not change when the ensemble size is varied within the range of what may currently be feasible in global applications.

We can extract more information from the data in Table 2 by assuming that statistical errors are negligible for our largest (128-member) ensemble. In other words, we assume that residual errors (not due to finite ensemble size) are approximately given by the estimation errors from the 128-member ensemble. This assumption allows us to isolate the statistical noise for smaller ensemble sizes N by subtracting the mean-square error (MSE) of the 128-member ensemble from the total MSE for N members. Figure 10 shows these presumably statistical errors as a function of the ensemble size. From theory we expect that the statistical MSE decreases in proportion to the inverse of the ensemble size. A regression analysis produces a slope of -1.2 for the 1D-EnKF and -1.1 for the 3D-EnKF. The 95% confidence intervals for the slope $[(-1.5, -1.0)$ for the 1D-EnKF and $(-1.2, -1.0)$ for the 3D-EnKF] include -1 , which is consistent with our assumption that the statistical error is negligible for 128 ensemble members.

Using our rough estimate of the residual error from the 128-member ensemble, we can now see from Fig. 10 that for the 1D-EnKF with 12 ensemble members the statistical MSEs are more than one order of magnitude less than residual errors. This result is not surprising because in each independent catchment we only have three state variables and a correspondingly small number of effective degrees of freedom. The result is also consistent with a previous result that the 1D-EnKF

TABLE 2. Rmse of root zone moisture content ($100 \text{ m}^3 \text{ m}^{-3}$) for different ensemble sizes N .

N	4	8	16	32	64	128
1D-EnKF	2.00	1.92	1.88	1.88	1.87	1.87
3D-EnKF	1.69	1.56	1.49	1.46	1.44	1.43

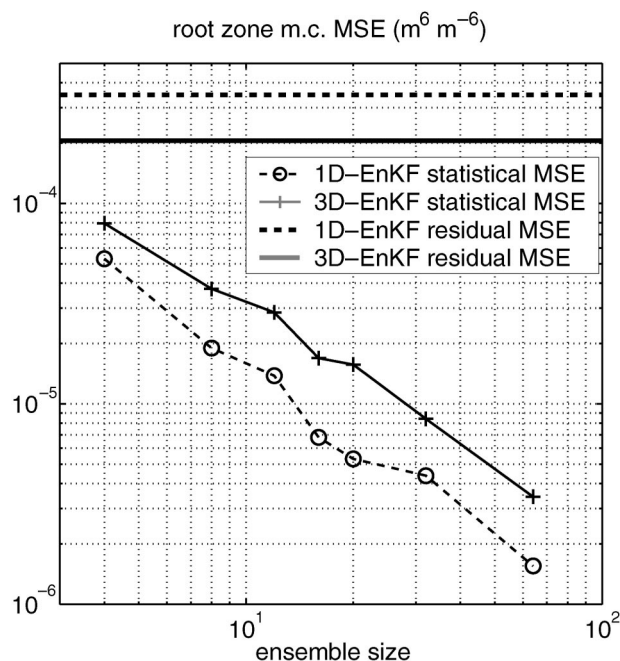


FIG. 10. Statistical and residual part of the time- and domain-average mean-square error (MSE) of root zone soil moisture. Residual MSEs are approximated by errors from the 128-member ensemble.

with four ensemble members performs similar to an extended Kalman filter (Reichle et al. 2002b). As expected, the statistical errors of the 3D-EnKF constitute a larger fraction of the total errors than in the 1D-EnKF. Nevertheless, total errors are always smaller in the 3D-EnKF and we can tolerate a larger share of statistical noise. In summary, very modest ensemble sizes on the order of 10 are acceptable for both filters.

There is of course a price to pay for the improved estimation accuracy of the 3D-EnKF. The additional cost of the 3D-EnKF depends strongly on the total volume of assimilated data, their distribution in time and space, the ensemble size, and the relative length scales of error correlations, resolution, and domain size. For our experiment, the 3D-EnKF requires about 1.6 times more CPU time than the 1D-EnKF primarily because of the additional expense of generating correlated error fields. In turn, the 1D-EnKF (with 12 ensemble members) requires about 12 times more CPU time than a straight model simulation without assimilation (prior). Relative to the prior, the computational cost is therefore 12 for the 1D-EnKF and 20 for the 3D-EnKF. This compares to a reduction of root zone soil moisture errors by $0.017 \text{ m}^3 \text{ m}^{-3}$ or 47% ($0.021 \text{ m}^3 \text{ m}^{-3}$ or 58%) for the 1D-EnKF (3D-EnKF) over the prior. The relative gain of using the 1D-EnKF over the prior is therefore larger than the relative gain of using the 3D-EnKF over the 1D-EnKF. Nevertheless, the additional computational expense of the 3D-EnKF may be justified depending on its specific application and appropriate measures of performance.

5. Summary and conclusions

We have investigated the importance of horizontal error correlations in background (i.e., model forecast) fields on soil moisture estimation by comparing the performance of one- and three-dimensional versions of the EnKF in a twin experiment. The main source of background error in our twin experiment was a difference in precipitation forcing, with the “true” model solution using gauge-based precipitation, and the prior and assimilation experiments using reanalysis precipitation. Assimilated synthetic surface soil moisture retrievals follow the spatiotemporal pattern of SMMR.

Both the 1D- and the 3D-EnKF produce satisfactory estimates of soil moisture that improve over the prior (no assimilation) estimates. Our findings also indicate that the 3D-EnKF produces more accurate soil moisture estimates than the 1D-EnKF. The most critical factor in determining the relative performance of the 1D- and 3D-EnKF is the data volume available for assimilation. The performance advantage of the 3D-EnKF is rooted in its ability to propagate observation information from observed to unobserved locations. In the least-observed catchments, soil moisture estimation errors are largest in both filters, and the 3D-EnKF has its strongest performance advantage. In the well-measured catchments, soil moisture estimation errors are generally small, and there is no difference in performance between the 1D- and the 3D-EnKF.

Further experiments in which the overall volume of assimilated data was reduced show that when very few data are available, the difference between the 1D- and 3D-EnKF is small because the impact of the few data on the estimates is small (relative to no assimilation). When the data volume is large, the performance of the 1D- and 3D-EnKF is again closer, but the 3D-EnKF still outperforms the 1D-EnKF. For intermediate data volumes, the performance advantage of the 3D-EnKF is most pronounced. While our analysis is based on the historic SMMR instrument, current and future sensors such as AMSR-E or L-band sensors will likely produce greater data volumes through more frequent overpasses or higher spatial resolution. This, however, does not imply that the performance difference between the 1D- and 3D-EnKF will vanish. The 3D-EnKF will presumably remain valuable, for example, for soil moisture estimation in densely vegetated areas, for which soil moisture remote sensing remains highly uncertain or even impossible, and the data volume remains low.

The seasonal distribution of prior (no assimilation) soil moisture estimation errors exhibits a maximum in late spring and early summer, precisely when accurate soil moisture initial conditions are most needed for the initialization of (seasonal) summer precipitation forecasts. Fortunately, both filters consistently produce improved soil moisture estimates throughout the year. The performance advantage of the 3D-EnKF over the 1D-

EnKF, however, is greatest during the critical spring and early summer period.

The performance difference between the 1D- and 3D-EnKF is robust when the ensemble size is varied. The 3D-EnKF with just 4 ensemble members still produces more accurate soil moisture estimates than the 1D-EnKF with 128 ensemble members. The disadvantage of the 3D-EnKF is its increased complexity, which makes it harder to implement and requires increased computational resources. In our experiment, the 3D-EnKF is about 1.6 times more expensive than the 1D-EnKF for the same number of ensemble members. The relative cost of the two filters depends on the total number of assimilated data, their distribution in time and space, the ensemble size, and the relative length scales of horizontal error correlations, domain size, and horizontal resolution of the computational units.

A basic assumption of the research presented here is that the difference in two precipitation datasets is representative of actual errors in global precipitation observations. At least until far more accurate global precipitation data are available, we feel that our approach is a reasonable starting point. In the future, other avenues for the quantification of error statistics of precipitation and other inputs will have to be explored. From an assimilation perspective, precipitation datasets would ideally be available with such error statistics. In our implementation an ensemble of precipitation fields is generated simply by perturbing the prior (reanalysis) precipitation with a multiplicative lognormal factor. More sophisticated approaches could be designed around rain/no-rain flags or by sampling precipitation realizations from conditional climatological distributions (Crow 2003).

The 3D-EnKF can lead to an increase in soil moisture estimation accuracy over that generated with the 1D-EnKF. Whether or not the improvement in accuracy is worth the additional computational expense depends on the application at hand. For seasonal prediction the importance of soil moisture initial conditions and accuracy requirements are still a topic of active research. The ultimate test will be whether or not more accurate soil moisture initial conditions lead to more accurate seasonal forecasts.

Acknowledgments. This research was sponsored by NASA Grant NRA-00-OES-07 (GWEC) and the NASA Seasonal-to-Interannual Prediction Project. We would like to thank Sarith Mahanama and Aaron Berg for their support with the data, Tom Hamill and an anonymous reviewer for their helpful comments, and Paul Houser, Wenge Ni-Meister, and Jeffrey Walker for many discussions.

REFERENCES

- Boone, A., and Coauthors, 2004: The Rhône-Aggregation Land Surface Scheme intercomparison project: An overview. *J. Climate*, **17**, 187–208.

- Bowling, L., and Coauthors, 2003: Simulation of high-latitude hydrological processes in the Torne–Kalix basin: PILPS phase 2(e) 1: Experiment description and summary intercomparisons. *Global Planet. Change*, **38**, 1–30.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Crow, W. T., 2003: Correcting land surface model predictions for the impact of temporally sparse rainfall rate measurements using an ensemble Kalman filter and surface brightness temperature observations. *J. Hydrometeorol.*, **4**, 960–973.
- , and E. F. Wood, 2003: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.*, **26**, 137–149.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- De Jeu, R. A. M., 2003: Retrieval of land surface parameters using passive microwave remote sensing. Ph.D. dissertation, Vrije Universiteit Amsterdam, 122 pp.
- Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model, 2, Parameter estimation and model demonstration. *J. Geophys. Res.*, **105**, 24 823–24 838.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.
- Gibson, J., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA description. ECMWF Re-Analysis Project Report Series 1, 72 pp. [Available from the European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, United Kingdom.]
- GuilleVIC, P., R. D. Koster, M. J. Suarez, L. Bounoua, G. J. Collatz, S. O. Los, and S. P. P. Mahanama, 2002: Influence of the interannual variability of vegetation on the surface energy balance—A global sensitivity study. *J. Hydrometeorol.*, **3**, 617–629.
- Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.
- , J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.
- Heemink, A. W., M. Verlaan, and A. J. Seegers, 2001: Variance reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, **129**, 1718–1728.
- Higgins, R. W., W. Shi, E. Yarosh, and R. Joyce, 2000: Improved United States Precipitation Quality Control System and Analysis. *NCEP/Climate Prediction Center Atlas*, No. 7, National Oceanic and Atmospheric Administration, 40 pp. [Available online at <http://www.cpc.ncep.noaa.gov/research-papers/ncep-cpc-atlas/7/index.html>.]
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- , and —, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.
- Keppenne, C. L., and M. M. Rienecker, 2002: Initial testing of a massively parallel Ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.*, **130**, 2951–2965.
- Koster, R. D., and M. J. Suarez, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeorol.*, **4**, 408–423.
- , —, A. Ducharme, M. Stieglitz, and P. Kumar, 2000a: A catchment-based approach to modeling land surface processes in a

- general circulation model, 1, Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822.
- , —, and M. Heiser, 2000b: Variance and predictability of precipitation at seasonal-to-interannual timescales. *J. Hydrometeor.*, **1**, 26–46.
- Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne, 2002: Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resour. Res.*, **38**, 1299, doi:10.1029/2001WR001114.
- Owe, M., R. de Jeu, and J. Walker, 2001: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1643–1654.
- Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.
- Reichle, R. H., D. Entekhabi, and D. B. McLaughlin, 2001: Downscaling of radiobrightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. *Water Resour. Res.*, **37**, 2353–2364.
- , D. B. McLaughlin, and D. Entekhabi, 2002a: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103–114.
- , J. P. Walker, R. D. Koster, and P. R. Houser, 2002b: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeor.*, **3**, 728–740.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Walker, J. P., and P. R. Houser, 2001: A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *J. Geophys. Res.*, **106**, 11 761–11 774.